Supplementary materials for

Direct Identification of Fish Species by Surface Molecular Transferring

Mingke Shao, Hongyan Bi*

College of Food Science and Engineering, Shanghai Ocean University, Hucheng Ring Road 999, Pudong New District, 201306 Shanghai, China

* To whom correspondence should be addressed. E-mail address: hybi@shou.edu.cn

E-mail address for the other authors: shaomingke_1015@163.com

S1. Photos and information on the analyzed fish samples

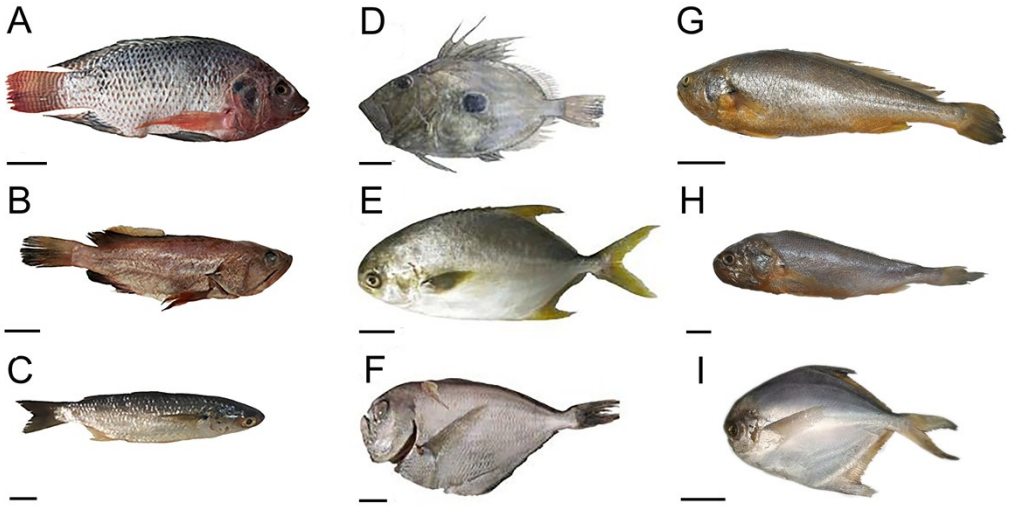

Fig. S1. Photos of fishes analyzed in the present study: (A) *Oreochromis mossambicus* (B) *Epinephelus rivulatus* (C) *Mugil cephalus*; (D) *Zeus faber* (E) *Trachinotus ovatus* (F) *Brama japonica* (G) *Larimichthys crocea* (H) *Larimichthys polyactis* (I) *Pampus argenteus.* Scale bar in each photo represents 1 cm.

Table S1. List of the scientific classification of fishes analyzed in the study. The classification of fishes refers to *https://www.fishbase.de/*.

Chinese name	**English common name**	**Binomial name (Scientific name)**	**Abbreviation**	**Scientific classification**
Japanese Wufang	Brama japonica	*Brama japonica*	BJ	Actinopterygii (class) > Perciformes (order) > Bramidae (family) > Brama (genus) > B. brama (species)
Baichang Fish	Silver pomfret; White pomfret	*Pampus argenteus*	PA	Actinopterygii (class) > Scombriformes(order) > Stromateidae（family) > Pampus (genus) > P. argenteus (species)
Haifang (commonly called:	Zeus faber Linnaeu; John Dory;	*Zeus faber*	ZF	Actinopterygii (class) > Zeiformes (order) > Zeidae (family) >

Yueliang Fish)	target perch			Zeus (genus) > Z. faber (species)
Luofei Fish	Mozambique tilapia	*Oreochromis mossambicus*	OM	Actinopteri (class) > Cichliformes (order) > Cichlidae (family) > Oreochromis (genus) > O.mossambicus (species)
Xiaozhai Fish	Flathead grey mullet	*Mugil cephalus*	MC	Actinopterygii (class) > Mugiliformes (order) Mugilidae (family) > Mugil (genus) > M. cephalus (species)
Shiban Fish	Halfmoon grouper	*Epinephelus rivulatus*	ER	Actinopterygii (class) > Perciformes (order) > Serranidae (family) > Epinephelus (genus) > E. rivulatus (species)
Dahuang Fish	Large yellow croaker	*Larimichthys crocea*	LC	Actinopterygii (class) > Perciformes (order) > Sciaenidae (family) > Larimichthys (genus) > L. crocea (species)
Xiaohuang Fish	Small yellow croaker	*Larimichthys polyactis*	LP	Actinopterygii (class) > Perciformes (order) > Sciaenidae (family) > Larimichthys (genus) > L. polyactis (species)
Jinchang Fish	Pompano	*Trachinotus ovatus*	TO	Actinopterygii (class) > Perciformes (order) > Carangidae (family) > Trachinotus (genus) > T. ovatus (species)

Table S2. Size and weight of fish used in the present study.

Fish species	**No. of fish samples**	**Size** / cm	**Weight** / g
Brama japonica	n = 6	18.9 ± 0.9	113.0 ± 11.4
Pampus argenteus	n = 6	29.8 ± 1.2	292.5 ± 12.1
Zeus faber	n = 6	22.2 ± 1.8	117.5 ± 0.5
Oreochromis mossambicus	n = 6	26.6 ± 0.7	245.0 ± 24.5
Epinephelus rivulatus	n = 6	23.8 ± 0.9	128.3 ± 9.4
Mugil cephalus	n = 6	17.9 ± 0.2	30.8 ± 4.5
Larimichthys crocea	n = 6	32.2 ± 1.1	425.8 ± 17.4
Larimichthys polyactis	n = 6	17.3 ± 0.4	53.3 ± 14.3
Trachinotus ovatus	n = 6	23.8 ± 0.9	24.8 ± 5.0

S2. Influence of matrix to mass spectral quality

Not all peptides can be observed per matrix, it depends on the nature of the matrix.[1] The choice of matrix has a great influence on the recorded mass spectrum in MALDI MS analysis since matrix plays an important role in ion yield and cleavage efficiency during the desorption/ionization of proteins from the crystal matrix. CHCA and SA are widely used matrices for the ionization of peptides or protein.[1] SA matrix has also been reported to be cable of increasing the heavy masses, while the signal loss for low m/z was relatively large.[2] To obtain better mass spectra, in this study, the fish samples of ER, MC and OM were analyzed by using CHCA and SA respectively. Fig. S2 shows the mass spectra of ER by using SA and CHCA respectively. It can be found when using CHCA matrix for the detection of ER, that the intensity of signal at m/z near 1370, 3690, 5600, and 8489 can be enhanced, however, the resolution and signal intensity of the CHCA matrix is relatively low at m/z larger than 10 000 Da. Table S3 compares the numbers of peaks detected under the assistance of SA and CHCA for fish species ER and OM. The number of detected mass peaks is 78 in the m/z range of 800-20 000 when SA matrix was used, while the number is 57 when CHCA matrix was used, showing that SA matrix can assist to obtain more mass spectral peaks. As shown in Figure S2, when CHCA and SA were used for the detection of fish OM, the findings keep consistency. In the present study, SA matrix was utilized during MS analysis to obtain relatively good mass signals.

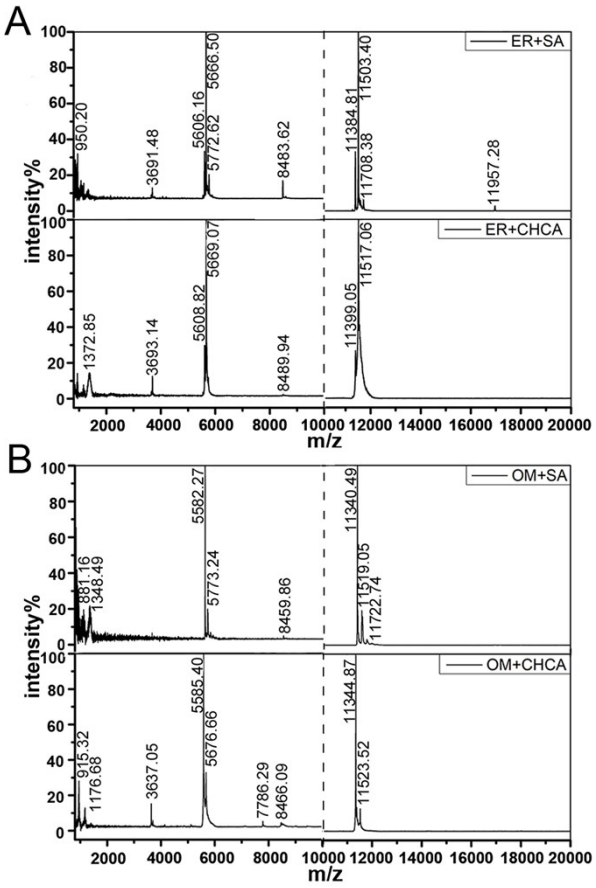

Fig. S2. (A) Representative MALDI mass spectra for materials on the surface of muscle tissue of fish ER under SA and CHCA matrices. The detection range is set as 800 - 20 000. (B) Representative MALDI mass spectra for materials on the surface of muscle tissue of fish OM under SA and CHCA. The detection range is set as 800 - 10 000, and 10 000 - 20 000, respectively. ER and OM in the inlet indicate the fish species *Epinephelus rivulatus* and *Oreocmis mossambicus* respectively. During the experiment, fish muscle tissues were cut to expose a cross section, and the bio-materials were transferred onto the MALDI plate surface by manually binding two surfaces for 2 to 5 seconds.

Table S3. Comparison of matrix for the MALDI MS detection of fishes. Fish species ER and OM were analyzed. SA matrix and CHCA matrix were used for their MALDI MS detection.

Fish species	*Epinephelus rivulatus*		*Oreochromis mossambicus*	
Detect mass range	800 - 2 0000			
Matrix	SA	CHCA	SA	CHCA
No. of Peaks with S/N≥3	78	57	71	61
Position of peaks (m/z) with S/N≥10	3691; 8483; 5606; 11384; 5666; 11503; 5772; 11708; 11957;	1372; 8489; 3696; 11399; 5608; 11517;	881; 1348; 5582; 5773; 8459; 11340; 11519;	951; 3637; 5585; 5675; 7786; 8466; 11344; 11523;

S3. Identification of fish by analyzing its skin

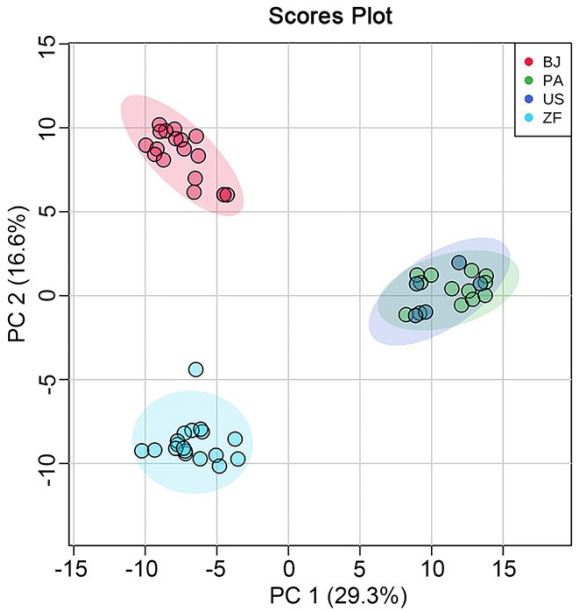

Fig. S3. The scores plot for the samples of reference fish and unknown fish. The fishes were analyzed by their skin samples. PA: *Pampus argenteus*, BJ: *Brama japonica*, ZF: *Zeus faber* and US: unknown fish.

S4. Ions obtained by molecularly imprinting method

Table S4. Positions of prominent peaks (with signal-to-noise ratio (S/N) ≥10) of mass spectra for fish LP, LC, OM, MC and ER, respectively. The fish muscle tissue was cut to expose a fresh cross section, and the bio-materials were transferred onto the surface of MALDI plate by manually pressing the cross section of fish muscle tissue onto the surface of the plate for 2 - 5 seconds.

Analyzed fish species	*Larimichthys polyactis*	*Larimichthys crocea*	*Oreochromis mossambicus*	*Mugil cephalus*	*Epinephelus rivulatus*
Position of peaks (m/z) with S/N≥10	908; 1500; 1518; 1558; 1566; 1586; 1586; 1597; 1604; 3363; 3790; 4871; 5688; 5743; 5590; 5690; 8385; 8438; 8539; 11380; 11405; 11449; 11586; 116774; 116920	813; 1061; 1078; 1104; 1149; 1533; 1580; 4089; 4847; 5753; 5777; 5798; 5870; 8394; 8402; 8473; 11393; 11472; 11509; 11598; 11734; 16792; 16808; 16939; 17017	1058; 1283; 3802; 5708; 5778; 5869; 8544; 11418; 11620; 11827; 11856; 11923; 11950; 11987; 12002; 12128; 12148; 12173	1179; 1303; 1497; 3756; 3806; 4052; 4381; 4664; 4863; 5555; 5631; 5710; 6199; 7851; 8325; 8492; 8508; 11268; 11414; 11619; 11960; 17006	1148; 1269; 3851; 3939; 4220; 4349; 5606; 5666; 5781; 5882; 8483; 8554; 11402; 11503; 11564; 11773; 17107

S5. Identification of fish by directly analyzing its muscle tissue

To apply the directly developed method for the identification of unknown fish, we endeavored to build a PCA score plot as a reference with the mass spectral datasets of five known fish species, and then determined a fish sample with unknown fish species by comparing the mass spectra of the unknown sample with the datasets of known samples, as shown in Fig. S4**.**

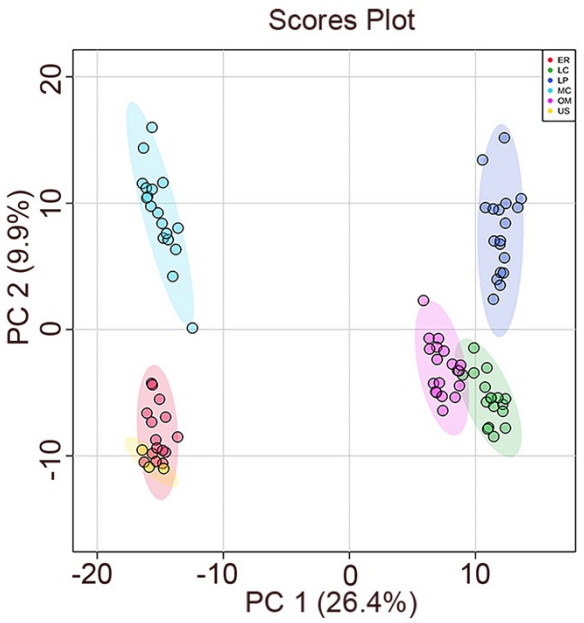

Fig. S4. The scores plot for the samples of reference fish and unknown fish. LP: *Larimichthys polyactis*, LC: *Larimichthys crocea*, OM: *Oreochromis mossambicus*, MC: *Mugil cephalus*, ER: *Epinephelus rivulatus* and US: unknown sample. The samples were analyzed by their muscle tissues.

S6. Comparison of ions in different preparation run

Table S5. Comparison of independent measurement of bio-materials on the surface of the dorsal part of fish muscle tissue. During the experiment, three fish species under SA matrix with the detection range of 800 - 20 000 were analyzed. Six fish samples of each fish species were taken for performing the experiment, and each experiment was triplicated. Positions of peaks were analyzed by Bruker flexAnalysis software.

	Run 1	Run 2	Run 3
Analyzed fish species		*Epinephelus rivulatus*	
No. of Peaks (S/N ≥ 3)	99 ± 1 (n = 6)	100 ± 0 (n = 6)	100 ± 0 (n = 6)
Position of peaks (m/z ± SD) (S/N ≥ 10)	3716.19 ± 4.21 5707.78 ± 1.99 8543.53 ± 2.70 11584.08 ± 3.06 11463.83 ± 3.20 17078.73 ± 5.98	3722.11 ± 0.79 5711.86 ± 1.30 8549.44 ± 1.97 11590.94 ± 2.26 11471.45 ± 2.56 17088.15 ± 3.92	3720.23 ± 0.74 5709.63 ± 0.98 8545.54 ± 1.37 11586.75 ± 1.37 11468.10 ± 1.41 17081.29 ± 3.59
Analyzed fish species		*Mugil cephalus*	
No. of Peaks (S/N ≥ 3)	98 ±2 (n = 6)	95 ± 4 (n = 6)	92 ± 4 (n = 6)
Position of peaks (m/z ± SD) (S/N ≥ 10)	1407.41 ± 2.36 3940.11 ± 1.35 5639.31 ± 3.58 11445.38 ± 6.43 12144.85 ± 5.67 16999.20 ± 8.03	 3931.63 ± 5.23 5646.24 ± 5.52 8518.76 ± 8.89 11459.70 ± 9.80 12124.21 ± 9.03 17019.79 ± 11.64	1405.40 ± 1.25 3935.04 ± 2.11 5648.10 ± 2.79 8519.22 ± 3.55 11463.52 ± 5.28 12120.14 ± 9.22 17024.89 ± 9.06
Analyzed fish species		*Oreochromis mossambicus*	
No. of Peaks (S/N ≥ 3)	96 ± 3 (n = 6)	98 ± 2 (n = 6)	99 ± 1 (n = 6)
Position of peaks (m/z ± SD) (S/N ≥ 10)	3785.85 ± 0.62 5684.68 ± 1.17 7842.66 ± 1.72 11378.04 ± 1.94 15692.09 ± 2.56	3785.12 ± 0.65 5684.15 ± 0.67 7841.98 ± 1.16 11376.45 ± 1.13 15690.39 ± 1.48	3780.60 ± 2.17 5862.81 ± 3.45 7840.11 ± 0.92 11375.66 ± 6.65 15674.78 ± 1.82

S7. Calculation of similarity score

For pattern matching identification, the experimental spectrum obtained from imprinted materials of the same part was compared. The spectral similarity score between mass spectra (*i* and j) was calculated by the cosine correlation method with the following Equation,[3]

$$\cos = \frac{\vec{y}_i \cdot \vec{y}_j}{\left|\vec{Y}_i\right| \cdot \left|\vec{Y}_j\right|} = \frac{\sum_{k=1}^{l} y_{ik} y_{jk}}{\sqrt{\sum_{t=1}^{n_i} Y_{it}^2} \cdot \sqrt{\sum_{t=1}^{n_j} Y_{jt}^2}} \qquad \text{(Equation S1)}$$

where y is the normalized intensity of a peak appearing in both spectrum i and spectrum j (an identical peak), l is the number of identical peaks in the two spectra, Y is the normalized intensity of a peak appearing in a spectrum and the number of peaks in a spectrum. Only peaks with S/N ≥ 3 are considered. Peaks appearing in different spectra with Δ (m/z)/(m/z) ≥ 1 000 ppm were considered as identical peaks. A tolerance of 2 000 ppm was chosen since linear mode TOF analysis was used.

S8. Comparison of mass spectral signals of transferred materials on the white muscle tissues of fish located an identical part of a fish

In statistics, frequency distribution can show, in a graph or dataset, the frequency of occurrence of each possible outcome of a repeatable event.[4] The similarity scores represent the closeness of characteristics between two mass spectra. And the similarity scores between mass spectra can be measured by comparing the mass spectra where cosine correlation is a commonly used method.[5, 6] The frequency distribution of similarity score has been used for bacteria identification.[3]

To further compare the mass spectra of the materials from different parts of an identical fish sample, the averaged mass spectra of imprinted materials from three parts of identical fish samples were compared. The similarity scores were calculated by the cosine correlation with the equation listed in Equation S1. And the frequency distribution was illustrated, as shown in Fig. S5, the similarity score between any two mass spectra for the same part of fish ER is ≥ 0.98, the similarity score for those of fish OM is ≥ 0.96, showing that the similarity in a close part of a fish is very high.

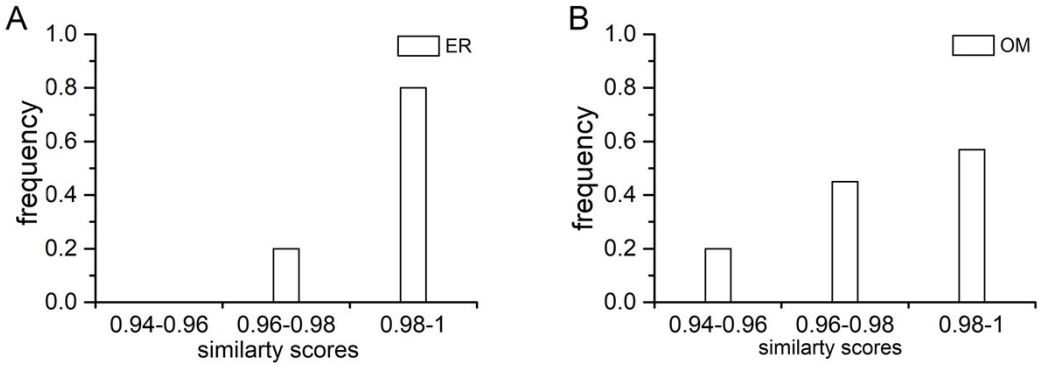

Fig. S5. Frequency distribution of similarity scores between the mass spectra for the muscle tissues from the same part of fish samples. (A) Mass spectra were obtained from the same part of ER; (B) Mass spectra were obtained from the same part of OM.

The parts for taking samples include the dorsal part (D), the ventral part (V) and the caudal part (C) of a fish.

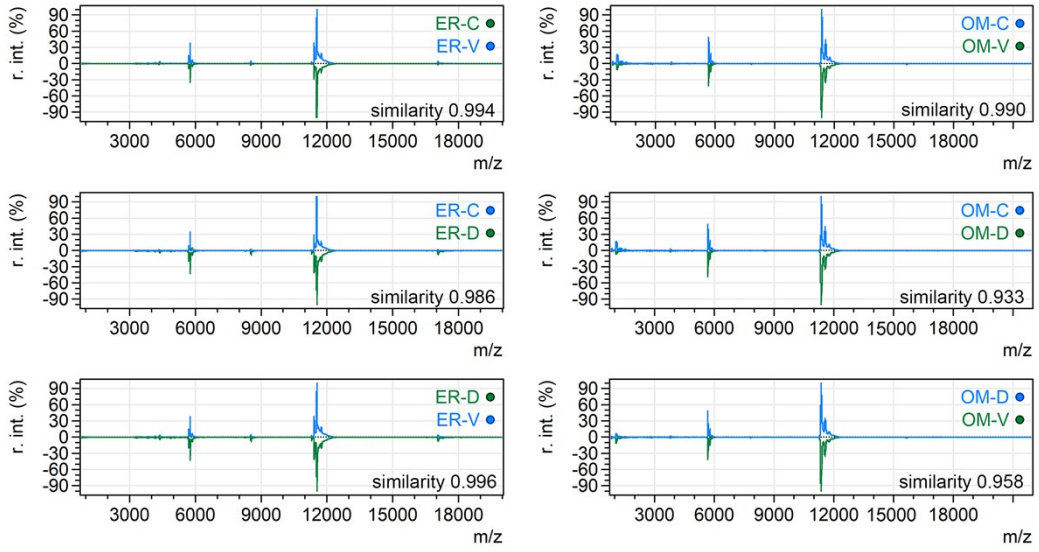

Fig. S6. Comparison of imprinted materials from three parts of identical fish samples. Spectral similarity score between mass spectra was calculated by the cosine correlation algorithm on *http://bacteriams.com/*. C, D, and V in the inlet are the caudal (C), dorsal (D), and ventral (V) part respectively. Peaks with S/N $\geqslant$ 3 are considered, with mass tolerance of 2 000 ppm. The analyzed fish samples are halfmoon grouper (*Epinephelus rivulatus*, ER) and mozambique tilapia (*Oreochromis mossambicus*, OM). Mass spectra peak picking was performed with the mMass Open Source Mass Spectrometry Tool on *http://www.mmass.org*.

References

1. Z. Liu and K. L. Schey, *Journal of the American Society for Mass Spectrometry*, 2005, **16**, 482-490.
2. L. H. C. Christine L. Gatlin-Bunai, William E. Cooke, Oliver J. Semmes, and Dariya I. Malyarenko, *Journal of Proteome Research*, 2007, **6**, 4517-4524.
3. Y. Zhu, L. Qiao, M. Prudent, A. Bondarenko, N. Gasilova, S. B. Möller, N. Lion, H. Pick, T. Gong and Z. Chen, P. Yang, L. T. Lovey and H. Girault, *Chem. Sci*, 2016, DOI: 10.1039.C1035SC04919A
4. S. Manikandan, *J Pharmacol Pharmacother.*, 2011, **2**, 54-56.
5. I. Koo, S. Kim and X. Zhang, *J Chromatogr A*, 2013, **1298**, 132-138.
6. Y. Yang, Y. Lin, Z. Chen, T. Gong, P. Yang, H. Girault, B. Liu and L. Qiao, *Anal Chem*, 2017, **89**, 12556-12561.